United States
Department of
Agriculture

Forest Service

**Rocky Mountain
Forest and Range
Experiment Station**

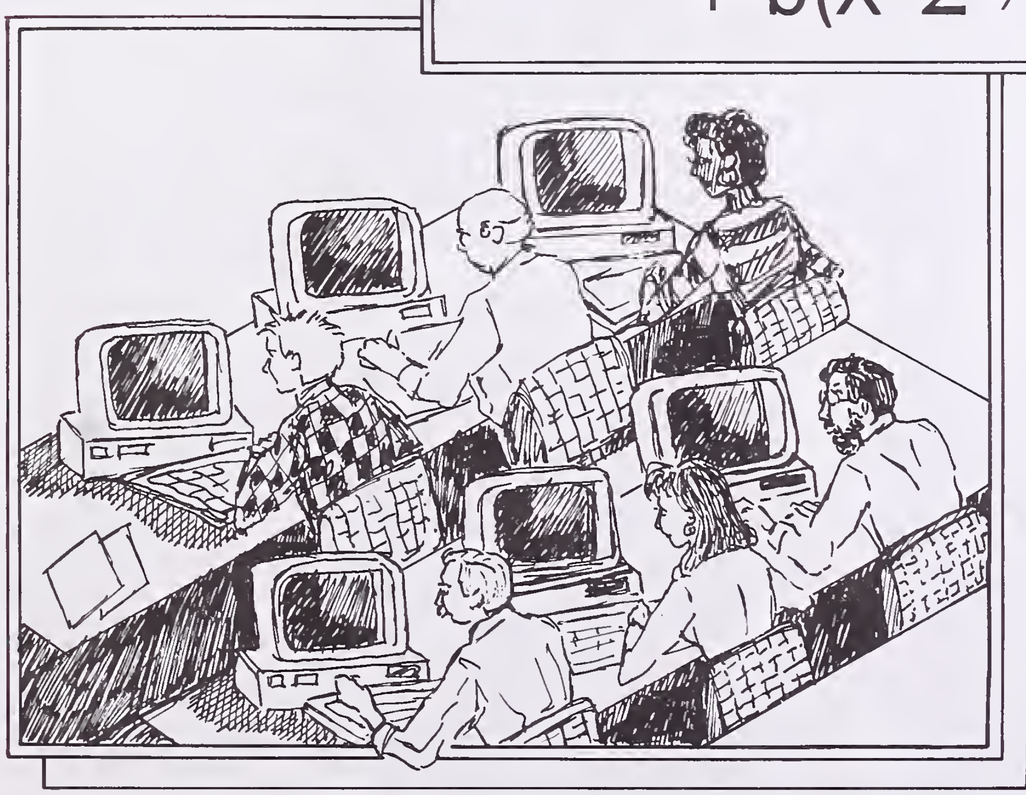
Fort Collins,
Colorado 80526

**Research Paper
RM-302**

ONEPHASE: A Simulation Program to Compare 1-Phase Sampling Strategies

Glen E. Brink and Hans T. Schreuder

$$\hat{Y}gr = \Sigma y/\pi + a(N - \Sigma 1/\pi) + b(X - \Sigma x/\pi)$$



USDA Forest Service
Research Paper RM-302

April 1992

ONEPHASE: A Simulation Program to Compare 1-Phase Sampling Strategies

Glen E. Brink, Computer Programmer Analyst
Rocky Mountain Forest and Range Experiment Station[1]

and

Hans T. Schreuder, Project Leader
Rocky Mountain Forest and Range Experiment Station

ABSTRACT

ONEPHASE is a computer simulation program primarily intended for use by students in Biometry or Forest Mensuration. Using real or artificial populations, it simulates the results of several inventory sampling techniques using several regression estimators and the Horvitz-Thompson estimator. Both volume estimates and variance estimates are generated and the results are displayed for comparison and analysis. Parameters, such as sample size, can be varied among runs and their influence examined. While the purpose of the paper is to provide a classroom tool and not necessarily to draw conclusions on management implications, managers may also find the program of use in weighing alternative inventory methods.

[1]*Headquarters is in Fort Collins, in cooperation with Colorado State University.*

ONEPHASE: A Simulation Program to Compare 1-Phase Sampling Strategies

Glen E. Brink and Hans T. Schreuder

Management Implications

The computer program used in the simulation study by Schreuder et al. (1990) compared traditional sampling strategies with model-based procedures and emphasized results of interest to managers who must select among various sampling techniques. It was modified in Schreuder and Ouyang (1992) to compare only design-based sampling designs with regression estimators and the Horvitz-Thompson estimator. ONEPHASE is a simplified version of that modified program and is available to managers for simulations on their own populations of interest and for classroom examination of sampling designs and estimators.

A word to students in such a classroom examination who may be new to sampling, statistics, and simulation: When the senior author was taking college mathematics courses, publishers of textbooks were just beginning to include answers to odd-numbered problems in the back of the book and computers were becoming available for producing answers rapidly and accurately. Both tended to enhance the perception that there was always one right answer. Introductory statistics was somewhat of a shock as ideas of sample variances impinged on that perception. Then knowledge of different variance formulas that yielded similar, but different, results obliterated it. This paper is intended to introduce the idea that there are different "answers" in forest inventory, and "the right one" is a matter of objectives, opinion, and debate.

Introduction

In computer simulation, we have the luxury of knowing "truth" either by treating large survey samples as populations or generating populations with known characteristics. Thus, we can simulate inventorying for some parameter, such as the total wood volume of the population, and compare the inventory estimates with "truth." This allows us to study the accuracy and precision of each technique relative to "truth" and examine the variability of the competing estimates.

Both Mackisack and Wood (1988, 1990) and Arvanitis and Reich (1989) developed related sampling technique computer programs that are useful in classroom instruction for studying the effects of sampling techniques. Both techniques rely on artificial populations that are generated by user-specified parameters or conditions. Schreuder and Ouyang (1992) used several "live" populations with different characteristics in their study. We use one of these populations for illustration. All are available with the ONEPHASE program for further exploration.

Snedecor and Cochran (1967) discussed the use of regression in sampling, particularly with reference to the measurement of more than one attribute or variable in a sample. In such cases, one variable is often dependent on another, and regression can be used to predict the dependent variable (usually denoted as y) from the independent variable (x). In our data sets, the parameter of interest is volume. Volume is difficult, and therefore costly, to measure. However, many investigators have observed a consistent relationship between volume and d^2h, where d is the diameter at breast height and h is height, both easily obtained measurements. We assume tree diameter and height are known for the entire population, but measurement of volume is on only the sample units drawn by simulation. The regression estimators then predict the volume of the entire population. While the examples are all based on the volume/d^2h relationship, ONEPHASE can handle any dependent/independent variable relationship. The efficiency of sampling strategies depends on the underlying relationship between such variables.

Portability

ONEPHASE is written in FORTRAN 77, with two extensions that are widely available in current compilers; viz., variable names longer than 6 or 7 characters and the STRUCTURE/RECORD statements that allow for a more structured programming style. ONEPHASE was programmed for use on personal computers (IBM PC-compatible) and runs successfully when compiled in Microsoft Fortran 5.0 and in MicroWay NDP Fortran-486; it would probably compile with little or no difficulty on any compiler supporting the STRUCTURE/RECORD statements. The program could be converted to strict FORTRAN 77 standards by the tedious, but straightforward, processes of editing variable names and utilizing COMMON blocks to replace the STRUCTURE/RECORD statements. Copies of the Microsoft executable version for the PC/DOS environment or of the source code may be requested. A READ.ME file provides a users guide. A test data set, input parameter file, and corresponding output are also provided on the diskette. The DOS version of ONEPHASE allows for a population size of 2,000 in the input file. Time required is dependent on the number of simulations and bootstrap iterations: the test run included on the distribution diskette is small (40 iterations, 200 bootstrap samples) and required 10.5 minutes on a microcomputer with a 486

processor; the sample output shown later is from a much larger run (10,000 iterations, 200 bootstrap samples) and took 13 hours to complete. The distribution version of ONEPHASE will run without a coprocessor, but will be slower.

Simulating Randomness

The heart of any simulation that depends on "random" events or selections is the random number generator. Since a generator that always produces truly random sequences has not yet been developed, nor is very desirable because replication of experimental results would be impossible, simulators have relied on pseudorandom number generators that must be "seeded" to produce a series of random numbers. In the 1970's and early 1980's as we worked on time-sharing mainframes, we had great confidence in the capabilities of the generators we utilized. There were occasional rumblings from the theorists about "bad" generators, and we were cautious to limit our use to those from reputable statistical libraries or packages rather than those provided by manufacturers of compilers.

As our work migrated during the mid to late 1980's to the increasingly powerful PC environment, we found ourselves in search of portable, nonproprietary pseudorandom number generators. We found one that was simple to implement in FORTRAN, validated it with some simple tests for "randomness" (runs up and runs down) and uniformity (chi-square or Kolmogorov-Smirnov), and proceeded with our simulation study. Our results were counterintuitive—sometimes far too "good" and most often, alarmingly "bad." The underlying statistical theories were checked and rechecked; the FORTRAN code was examined and results were replicated by independent code. When we returned to validating the pseudorandom number generator, we found that we had fallen into the cyclic pattern trap so prevalent in PC random number generators. Our earlier tests had behaved satisfactorily when generating hundreds of thousands of pseudorandom numbers. Unfortunately, we were now generating millions of them (this is primarily due to the use of bootstrap methods for variance estimation); reapplying the basic tests disclosed that we had indeed reached the cyclic limit, and were regenerating the same sequences of "random" numbers.

A search of the literature revealed increasing interest among theorists that still has not peaked. L'Ecuyer (1988) discussed flawed generators and "periods that are too short to be used safely for serious applications." He proposed a portable generator for 32-bit computers which required that we work in double precision in FORTRAN; we found it to be somewhat slow because of the double precision (Microway's NDP-486 compiler handles the 32-bit data transfers more effectively, so the slowdown was not as perceptible), but it did pass our basic tests even when we generated millions of random numbers. The generator by Kahaner et al. (1988) was recommended to us at about this same time and it also passed all of our basic tests. As it was somewhat faster, we selected it as the generator for the simulation study.

L'Ecuyer's (1990) more recent comments indicate that caution is still the byword; but, in the realms of interest to us, these generators are probably viable. This brief discussion is not to explore the vagaries of pseudorandom number generators, but merely to alert the user to the fact that the choice of a generator can adversely affect the results of a simulation. We believe that the one chosen for the simulation study performs satisfactorily for the quantity of random numbers we required. We recognize, however, that both of the generators we examined probably have identifiable periods or cycles somewhere beyond our current requirements, and that if we were to require billions of random numbers, we would have to retest our generators and, perhaps, search again for an acceptable generator.

Simulating Sampling Selection Methods

ONEPHASE simulates five different sampling selection methods: (1) Restricted Simple Random Sampling (RSRS); (2) Sampling Proportional to Size from Cumulated X's (SPSCX); (3) Sampling with Probability Proportional to Size (SPPS); (4) Stratified Simple Random Sampling (STSRS); and (5) User-Defined Strata (USERDEF). Two variations of 3) are also simulated; SPPS with a U-distribution and Modified SPPS (SPPSU and SPPSMOD). All methods are described below. For each method, ONEPHASE computes 5 separate estimates of volume, four regression and one ratio estimator; these are described later. For each estimate, 6 different variances are computed. As mentioned before, we assume for illustration purposes that we are interested in volume and that the covariate used in both sample selection and estimation is d^2h. A small segment of the complete example output (which is displayed later) below shows the 5 volume estimators down the side and the 6 variances across the top. The symbols displayed in the output will be referenced in the subsequent sections, which discuss sampling methods, volume estimators, and variances.

	VZ	VZ1	VT	VT1	VB	VJ
$\hat{Y}_{gr}$	92.4	94.8	93.9	88.3	79.5	103.2
$\hat{Y}_{pi}$	78.9	81.0	80.5	77.7	91.8	114.1
$\hat{Y}_{piwr}$	111.5	114.4	113.3	106.5	95.9	104.0
$\hat{Y}_{wr}$	111.5	114.4	113.3	106.5	95.9	104.0
$\hat{Y}_{ht}$	98.4					

Sample Selection Methods

Based on the review of literature, in particular Schreuder and Ouyang (1992), the following sample selection methods are compared:

1. RSRS (Restricted Simple Random Sampling)—Units are selected randomly from the entire population, and the mean and variance of the sample are compared to the population mean and variance. If either is not within the user-specified range, the sample is rejected

and another sample is selected. By making the specified range sufficiently large, the user may employ this method for Simple Random Sampling (SRS), the most basic sample selection method. We employ SRS in the example output by specifying a wide range (± 50%) around the population mean and variance.

2. SPSCX (Sampling Proportional to Size from Cumulated X's)—As discussed in Schreuder et al. (1990), the population is ordered by ascending values of x and stratified such that all strata have identical (or as nearly identical as possible) sums of the x^k values. One unit is then randomly selected from each stratum for the samples. The user specifies the value(s) of k to be used.

3. SPPS (Sampling with Probability Proportional to Size)—As discussed in Li et al. (1992), the population is ordered and stratified based on x^k such that certain properties of the strata are optimized for sampling purposes. The sample unit (or units) for each stratum is selected such that the larger units are more likely to be picked. Randomness still plays an important part as the x^k values are accumulated in random order and the first unit exceeding a randomly derived target value is the one selected. It is possible in a given stratum for a unit that is much larger than its neighbors to be selected with certainty on every sample. ONEPHASE accommodates such a case, but the user is advised to be aware of such anomalies in the population. The user specifies the value(s) of k and the number(s) of sample units to be selected from each stratum. There are also two special cases of SPPS sampling considered in ONEPHASE:

3a). SPPSU (U-distribution)—The probabilities of selection are computed such that the smallest and largest units are likely to be selected, while the middle units are less likely to be picked. The user has no control over ONEPHASE's handling of this sampling technique (k is always 1.0). This technique is more fully detailed in Schreuder and Ouyang (1992).

3b). SPPSMOD (Modified Sampling with Probability Proportional to Size)—The only difference between this sampling technique and the standard SPPS method is the guarantee that 1/4 of the sample units will be selected from the smallest units in the population. The subpopulation of small units set aside is 1/2 the sample, and the guaranteed selection of small units is done by simple random sampling from that subpopulation. The remainder of the population is then sampled by SPPS. The user may specify values of k in this technique.

4. STSRS (Stratified Simple Random Sampling)—The population is optimally stratified as in SPPS above, but sampling is by simple random sampling in each stratum. That makes this method similar to SPSCX except that the strata with the smaller x-values should be smaller and those with the larger x-values larger.

5. USERDEF (User-Defined Strata)—The user specifies the number of strata, the number of units to be assigned to each stratum, and the number of sample units to be drawn from each stratum by simple random sampling. For example, we chose to have 3 strata with the first and last strata containing a certain number of the smallest and largest units, respectively; 1/4 of the sample was drawn from each of them and the other 1/2 of the sample was drawn from the large middle stratum, thus guaranteeing a given proportion of small and large units in the sample.

Estimators

Five estimators are utilized in ONEPHASE, with the first four being regression estimators. The first estimate of volume displayed in the output segment shown earlier is the generalized regression estimator, $\hat{Y}_{gr}$ (Särndal 1980, 1982):

$$\hat{Y}_{gr} = \sum_{i=1}^{n} y_i/\pi_i + a_{gr}\,(N - \sum_{i=1}^{n} 1/\pi_i) + b_{gr}\,(X - \sum_{i=1}^{n} x_i/\pi_i) \qquad [1]$$

with estimated regression coefficients

$$a_{gr} = [\sum_{i=1}^{n} y_i/(\pi_i v_i) - b_{gr} \sum_{i=1}^{n} x_i/(\pi_i v_i)] / \sum_{i=1}^{n} 1/(\pi_i v_i) \qquad [1a]$$

and

$$b_{gr} = \frac{\sum_{i=1}^{n} (x_i - \tilde{x}) y_i/(\pi_i v_i)}{\sum_{i=1}^{n} (x_i - \tilde{x})^2/(\pi_i v_i)} \qquad [1b]$$

where

n = sample size
Y_i = dependent variable for unit i
π_i = probability of selecting unit i (discussed below)
N = population size
x_i = independent variable for unit i

$$X = \sum_{i=1}^{n} x_i$$

v_i = variance for unit i, usually $v_i = x_i^k$ (k can be user-specified, we use 1.5)

$$\tilde{x} = \tilde{N}^{-1} \sum_{i=1}^{n} x_i/(\pi_i v_i)$$

$$\tilde{y} = \tilde{N}^{-1} \sum_{i=1}^{n} y_i/(\pi_i v_i)$$

$$\tilde{N} = \sum_{i=1}^{n} 1/(\pi_i v_i)$$

The probability of selection, π_i, is dependent on the sampling method:

$\pi_i \doteq n/N$ for RSRS
$= 1/n_s$ for SPSCX
$= n_s x_i^k / X_s$ for SPPS and SPPSMOD
$= n(x_i^* + 1/x_i^*)/X_s^*$ for SPPSU
$= n_s/N_s$ for STSRS and USERDEF

where

n_s = sample size drawn from a given stratum
N_s = the given stratum's size

$$X_s = \sum_{i=1}^{N_s} x_i$$

$$x_i^* = x_i / \text{median}(x)$$

$$X_s^* = \sum_{i=1}^{N} x^*_i + 1/x^*_i)$$

k = user-defined parameter as discussed in (3) SPSS above.

The second, third, and fourth estimators shown in the output segment are weighted regression estimators $\hat{Y}_{pi}$, $\hat{Y}_{piwr}$ and $\hat{Y}_{wr}$, respectively (Schreuder et al. 1990). All three are of the form

$$\hat{Y}_{wtregr} = Na_{wtregr} + b_{wtregr}X \quad [2]$$

with

$$a_{wtregr} = (\sum_{i=1}^{n} y_i/w_i - b_{wtregr} \sum_{i=1}^{n} x_i/w_i) / \sum_{i=1}^{n} 1/w_i) \quad [2a]$$

$$b_{wtregr} = \frac{\sum_{i=1}^{n} 1/w_i \sum_{i=1}^{n} x_i y_i/w_i - \sum_{i=1}^{n} y_i/w_i \sum_{i=1}^{n} x_i/w_i}{\sum_{i=1}^{n} 1/w_i \sum_{i=1}^{n} x_i^2/w_i - (\sum_{i=1}^{n} x_i/w_i)^2} \quad [2b]$$

where $w_i = \pi_i$ for $\hat{Y}_{pi}$,
$= \pi_i v_i$ for $\hat{Y}_{piwr}$, and
$= v_i$ for $\hat{Y}_{wr}$.

Note that $\hat{Y}_{pi}$ considers only the probability of selection, $\hat{Y}_{wr}$ considers only the estimated variance, but $\hat{Y}_{piwr}$ considers both the probabilities of selection (π_i) and variance weights ($v_i = x_i^k$) in estimating the regression coefficients and, in fact, its regression coefficients are the same as [1a] and [1b] above.

The fifth estimator shown in the output segment is the Horvitz-Thompson estimator, $\hat{Y}_{HT}$ (Cochran 1977):

$$\hat{Y}_{HT} = \sum_{i=1}^{n} y_i/\pi_i \quad [3]$$

Variance Estimators

The first two variance estimators displayed in the output segment above are VZ and VZ1, suggested by Ouyang et al. (1992):

$$v_{z1}(\hat{Y}_{regr}) = \sum_{i=1}^{n} (y_i - a - bx_i)^2/\pi_i^2$$

and

$$v_{z1}(\hat{Y}_{regr}) = [\sum_{i=1}^{n} (y_i - a - bx_i)^2/\pi_i^2] \, [n/(n-1)]$$

For the Horvitz-Thompson estimator, the jackknife variance estimator VZ is displayed and is calculated by:

$$v_z(\hat{Y}_{regr}) = [(N-n)/(Nn)] \, [1/(n-1)] \sum_{i=1}^{n} [(ny_i/\pi_i) - \hat{Y}_{regr}]^2$$

(Cumberland and Royall 1981).

Särndal (1980, 1982) proposed two variance estimators for $\hat{Y}_{gr}$: VT and VT1 in the output segment; note that for comparison purposes we have used these variance estimators for $\hat{Y}_{pi}$, $\hat{Y}_{piwr}$ and $\hat{Y}_{wr}$ as well.

$$v_t(\hat{Y}_{gr}) = \sum_{i<j}^{n} \frac{\pi_i \pi_j - \pi_i}{\pi_{ij}} (e_i/\pi_i - e_j/\pi_j)^2$$

$$v_{t1}(\hat{Y}_{gr}) = \sum_{i<j}^{n} \frac{\pi_i \pi_j - \pi_i}{\pi_{ij}} (é_i/\pi_i - é_j/\pi_j)^2$$

where π_{ij} = joint probabilities of selecting units i and j,

$$e_i = (y_i - \bar{y}) - b(x_i - \bar{x})$$

and

$$é_i = e_i - e_i\{ [(\hat{N}-N) \sum_{\ell=1}^{n} x_\ell^2/(v_\ell \pi_\ell) - (\hat{X}-X) \sum_{\ell=1}^{n} x_\ell/(\pi_\ell v_\ell)]/v_i$$

$$+ [-(\hat{N}-N) \sum_{\ell=1}^{n} x_\ell/(v_\ell \pi_\ell) + (\hat{X}-X) \sum_{\ell=1}^{n} 1/(\pi_\ell v_\ell)] x_i/v_i\}$$

$$* 1/\{ \sum_{\ell=1}^{n} x_\ell^2/(\pi_\ell v_\ell) \sum_{\ell=1}^{n} 1/(\pi_\ell v_\ell) - [\sum_{\ell=1}^{n} x_\ell/(\pi_\ell v_\ell)] \}$$

with $\hat{N} = \sum_{\ell=1}^{n} 1/\pi_\ell$ and $\hat{X} = \sum_{\ell=1}^{n} x_\ell/\pi_\ell$

Since $v_t(\hat{Y}_{gr})$ and $v_{t1}(\hat{Y}_{gr})$ vanish for sampling schemes where $\pi_{ij} = \pi_i \pi_j$ for units i and j such as for stratified sampling with one unit/stratum, alternatives to $v_t(\hat{Y}_{gr})$ and $v_{t1}(\hat{Y}_{gr})$ are

$$v_t(\hat{Y}_{gr}) = \sum_{i=1}^{n} \frac{(1-\pi_i)}{\pi_i^2} e_i^2 + \sum_{i \neq j}^{n} \frac{\pi_{ij} - \pi_i \pi_h}{\pi_i \pi_j \pi_{ij}} e_i e_j$$

$$v_{t1}(\hat{Y}_{gr}) = \sum_{i=1}^{n} \frac{(1-\pi_i)}{\pi_i^2} é_i^2 + \sum_{i \neq j}^{n} \frac{\pi_{ij} - \pi_i \pi_h}{\pi_i \pi_j \pi_{ij}} é_i é_j$$

which for $\pi_i \pi_j = \pi_{ij}$ reduces to

$$v_t(\hat{Y}_{gr}) = \sum_{i=1}^{n} (1-\pi_i) e_i^2/\pi_i^2$$

$$v_{t1}(\hat{Y}_{gr}) = \sum_{i=1}^{n} (1-\pi_i) é_i^2/\pi_i^2$$

(Schreuder and Ouyang 1992).

The last two variance estimators in the output segment are the bootstrap variance (VB) and the jackknife variance (VJ). Efron and Tibshirani (1991) present an excellent general discussion of the bootstrap for the nonstatistician and Zahl (1977) similarly describes the jackknife. Both are computed for only the first four

volume estimates $\hat{Y}_{gr}$, $\hat{Y}_{pi}$, $\hat{Y}_{piwr}$ and $\hat{Y}_{wr}$; in the following discussion, we let $\hat{Y}_{regr}$ represent any of the four.

In bootstrapping, from the sample of n units drawn from the population by any of the four methods, n_B bootstrap samples are selected by simple random sampling (SRS) with replacement. This generates n_B regression estimates $\hat{Y}_{regr}$. If these regression estimates are denoted generally by

$$\hat{Y}^{(b)}_{regr}, b = 1, \ldots, n_B$$

then we can compute a simple bootstrap variance estimate

$$v_B(\hat{Y}_{regr}) = \sum_{b=1}^{n_B} (\hat{Y}^{(b)}_{regr} - \hat{Y}^{(B)}_{regr})^2 / (n_B - 1)$$

where

$$\hat{Y}^{(B)}_{regr} = \sum_{b=1}^{n_B} \hat{Y}^{(b)}_{regr} / n_B$$

In jackknifing, as in bootstrapping, a sample of n units is drawn from the population by any of the four methods; then, one unit at a time is deleted from that sample and the regression estimate $\hat{Y}_{regr}$ is recomputed for the reduced sample size. If these n jackknife estimates are denoted in general by

$$\hat{Y}^{(j)}_{regr}, j = 1, \ldots, n$$

then the jackknife estimates (sometimes called pseudo-estimates or reduced estimates) are

$$\hat{Y}^{(J)}_{regr} = n\hat{Y}_{regr} - (n-1)\, \hat{Y}^{(j)}_{regr}$$

and the variance between these estimates is computed as

$$v_J(\hat{Y}_{regr}) = \sum_{j=1}^{n} (\hat{Y}^{(j)}_{regr} - \hat{Y}^{(J)}_{regr})^2 / (n(n-1))$$

where

$$Y^{(J)}_{regr} = \sum_{j=1}^{n} \hat{Y}^{(j)}_{regr} / n$$

Example Output

Included below are the results for one of the populations displayed in Schreuder and Ouyang (1992). Differences between the results displayed here and in that paper should alert the interested user to the fact that computers are not magically correct and that the computer programs themselves may yield similar, but different, results depending on programming language and style. Efforts were made to make the results comparable, but differences still persist. For example, starting even a single program with different random number generator seeds will not produce identical results. Thus, in the simplification of the earlier program to produce ONEPHASE, calls to the random number generator were performed in a slightly different order, resulting in small, but discernable, discrepancies.

Because our purpose is the comparing and contrasting of different methods, estimators, and variances instead of actual predictions of volumes, we chose to express the output in percentages relative to "truth." Both the simulation bias and the simulation standard error are displayed as a percentage of the total measured volume of the population. The average standard errors of estimates are stated as a percentage of the simulation standard errors, as the latter are our best measures of the actual standard errors. ONEPHASE first displays the input parameters and then displays tables of percentages as described above, plus the corresponding confidence intervals.

Output for Loblolly Pine Data Set

```
ONEPHASE (version 05-06-91   11:06a).                         Run: 15-MAY-91 15:46:58
Output file: LOB.OUT
Input file: D:\HANS\ONEPHASE\DATA\LOB81D1.SRT
Input format: (T11,D10.2,T1,D10.2)
No. simulations:                                  10000
Print frequency:                                  10000
No. bootstrap samples:                              200
Random number gen. seed:                         479233
Sample size:              20
t alpha =    2.101, deg. freedom =    18
Rejection criteria for RSRS (mean):              50.000
Rejection criteria for RSRS (variance):          50.000
''variance k'' for WR:      1.50
''stratum k'' for SPSCX:      1.00
''stratum k'' for SPSCX:      1.50
''stratum k'' for SPSCX:      0.50
''stratum k'', n(i) for SPPS:      1.00       1
''stratum k'', n(i) for SPPS:      1.50       1
''stratum k'', n(i) for SPPS:      1.00      20
n(i) for STSRS:        1
```

k for SPPSMOD: 1.00
Number of strata for "User-Defined Strata": 3
Strata sizes for "User-Defined Strata": 20 –1 20
Sample sizes for strata in "User Defined Strata": 5 10 5

N: 1801, Σx: 21573.4, Σy: 10700539.6 , No. Simulations: 10000

	BIAS %	STD ERR %	STD ERR FOR CLASSICAL, BOOTSTRAP & JACKKNIFE AS % OF SIM. STD ERR						CONFIDENCE INTERVAL PERCENTAGES					
			VZ	VZ1	VT	VT1	VB	VJ	VZ	VZ1	VT	VT1	VB	VJ
RSRS														
Ygr	100.2	4.6	92.4	94.8	93.9	88.3	79.5	103.2	86.2	87.0	86.7	91.0	93.4	89.6
Ypi	100.8	4.4	78.9	81.0	80.5	77.7	91.8	114.1	80.5	81.4	81.2	85.9	92.0	90.7
Ypiwr	100.3	3.8	111.5	114.4	113.3	106.5	95.9	104.0	87.2	87.8	87.6	92.3	93.4	93.9
Ywr	100.3	3.8	111.5	114.4	113.3	106.5	95.9	104.0	87.2	87.8	87.6	92.3	93.4	93.9
Yht	99.9	26.0	98.4						89.3					
SPSCX, k = 1.0														
Ygr	100.0	2.9	97.0	99.5	95.3	95.2	126.7	115.6	94.0	94.7	93.6	93.6	93.2	92.4
Ypi	100.1	2.9	111.6	114.5	100.9	100.8	140.4	140.3	95.9	96.3	94.5	94.5	92.0	97.4
Ypiwr	100.4	3.0	92.6	95.0	90.9	90.9	120.8	106.9	93.3	94.0	92.7	93.0	93.2	95.9
Ywr	99.7	2.9	95.8	98.3	93.8	93.8	124.7	122.1	93.5	94.2	93.2	93.3	93.2	95.9
Yht	100.0	3.3	123.9						96.6					
SPSCX, k = 1.5														
Ygr	100.0	3.6	71.8	73.7	70.3	70.7	101.3	163.2	84.6	85.6	83.6	83.2	93.5	90.9
Ypi	100.0	3.5	95.9	98.4	78.6	79.4	113.2	208.5	91.7	92.3	87.2	86.8	92.0	96.4
Ypiwr	100.5	3.7	69.3	71.1	67.8	68.2	97.7	143.3	83.5	84.5	82.3	82.3	93.5	93.8
Ywr	99.7	3.7	94.0	96.4	75.6	76.1	98.8	212.2	89.8	90.4	84.8	84.8	93.5	97.1
Yht	99.9	6.9	228.9						99.9					
SPSCX, k = 0.5														
Ygr	100.0	3.2	102.4	105.1	101.2	100.3	114.5	108.0	94.7	95.2	94.5	94.6	93.3	92.7
Ypi	100.2	3.0	102.7	105.4	101.9	100.5	131.8	120.3	94.2	94.7	94.1	94.3	92.1	95.8
Ypiwr	100.3	3.0	109.1	111.9	107.8	106.8	121.9	105.1	94.5	95.0	94.3	94.5	93.3	95.8
Ywr	100.1	3.0	103.1	105.8	102.1	101.1	119.6	108.4	94.1	94.7	93.9	94.1	93.3	95.4
Yht	100.0	3.4	307.1						100.0					
SPPS, k = 1.5, n(i) = 1														
Ygr	100.0	2.8	102.3	105.0	100.6	100.8	130.8	111.0	95.0	95.5	94.7	94.6	93.3	93.2
Ypi	100.0	2.8	109.7	112.6	104.2	104.6	142.6	127.6	95.8	96.2	95.1	95.1	91.9	97.2
Ypiwr	100.4	2.9	99.0	101.6	97.4	97.6	126.5	105.6	94.8	95.3	94.5	94.5	93.3	95.8
Ywr	99.6	2.9	95.6	98.1	94.4	94.7	124.3	116.5	93.2	93.8	92.9	93.0	93.3	95.9
Yht	100.0	3.3	153.1						99.4					
SPPS, k = 1.0, n(i) = 20														
Ygr	100.0	3.2	86.8	89.1	87.2	91.0	114.7	121.7	91.3	91.9	91.3	91.9	93.2	92.5
Ypi	100.7	3.9	78.6	80.7	70.8	76.9	101.9	122.5	89.8	90.4	87.4	89.2	92.3	96.4
Ypiwr	100.4	3.3	84.2	86.4	84.6	88.2	111.1	117.4	90.8	91.5	90.9	91.5	93.2	95.2
Ywr	99.9	3.4	81.1	83.2	80.8	86.3	106.4	121.9	90.0	90.7	89.6	90.9	93.2	95.1
Yht	100.0	3.8	100.3						94.0					
SPPSU														
Ygr	100.0	3.4	95.1	97.6	96.1	97.8	107.0	103.1	92.3	92.8	92.5	93.6	93.3	92.7
Ypi	100.2	3.6	89.5	91.9	90.8	92.7	109.7	109.5	90.9	91.7	91.7	92.9	92.0	95.0
Ypiwr	100.2	3.5	93.1	95.6	94.1	95.8	104.7	102.5	91.6	92.4	92.0	93.6	93.3	95.2
Ywr	99.2	3.8	85.2	87.4	85.9	87.5	96.3	104.7	89.0	89.9	89.2	90.6	93.3	93.8
Yht	100.0	14.6	98.7						94.6					
STSRS														
Ygr	100.0	3.2	102.7	105.4	101.5	100.7	114.7	108.4	95.1	95.5	94.8	95.0	93.5	92.9
Ypi	100.2	3.0	102.8	105.5	102.1	100.8	131.7	120.3	94.5	95.1	94.4	94.7	92.0	96.1
Ypiwr	100.3	3.0	109.5	112.3	108.2	107.3	122.1	106.0	95.0	95.5	94.8	95.1	93.5	96.0
Ywr	100.2	3.0	103.2	105.9	102.2	101.4	119.5	108.7	94.6	95.1	94.4	94.6	93.5	95.8
Yht	100.0	3.4	307.8						100.0					

SPPSMOD														
Ygr	100.0	3.5	98.8	101.3	97.8	97.7	102.9	107.2	93.5	94.0	93.2	93.2	93.3	91.6
Ypi	100.0	3.5	99.9	102.5	99.5	99.5	113.6	122.8	93.1	93.8	93.2	93.2	91.9	95.4
Ypiwr	100.2	3.4	103.4	106.1	102.4	102.3	107.6	105.7	93.6	94.3	93.4	93.5	93.3	95.6
Ywr	100.0	3.6	96.1	98.6	94.8	94.8	100.1	107.6	93.7	94.4	93.4	93.4	93.3	95.4
Yht	100.0	3.6	495.9						100.0					
USERDEF														
Ygr	99.8	5.0	88.2	90.5	91.5	89.4	73.3	112.8	81.5	82.2	82.0	86.8	93.4	89.2
Ypi	99.3	5.1	84.5	86.7	88.4	91.0	77.8	112.2	80.9	81.7	82.3	86.6	91.9	89.7
Ypiwr	100.1	4.8	90.2	92.5	93.6	91.5	74.9	111.5	81.1	81.8	81.7	86.7	93.4	92.1
Ywr	95.2	3.4	169.3	173.7	164.5	162.4	107.6	117.3	78.6	79.5	77.0	83.9	93.4	81.4
Yht	100.3	30.4	119.6						90.7					

Graphic Displays

Because the implications are not readily apparent in tabular output, it is suggested that users display ONEPHASE's results of interest in graphs such as figures 1 and 2, which were produced by a proprietary graphics software package. One can "see" that the bootstrap variance consistently underestimates the variance and the jackknife variance consistently overestimates it, but that one or the other behaves better in most cases than the other variance estimators. It is also apparent that the sampling method is critical if one chooses the Horvitz-Thompson estimator instead of one of the regression estimators. Such conclusions can be drawn from the tables, but the graphic display is more intuitive.

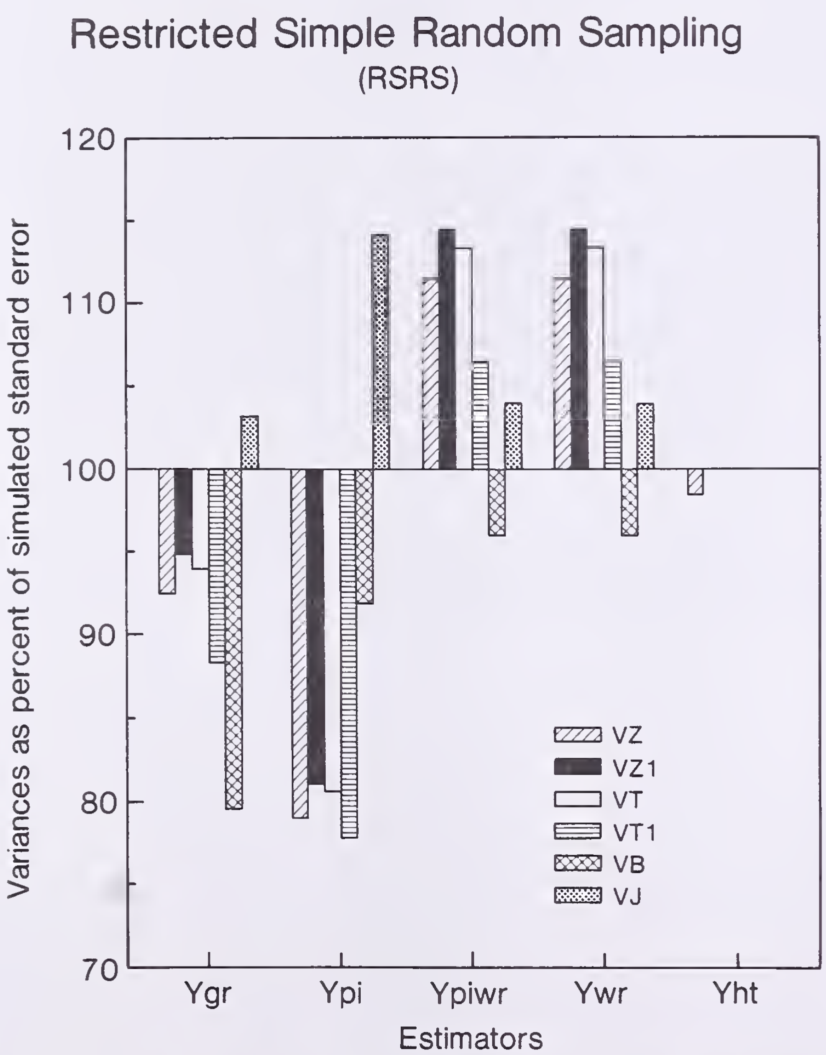


Figure 1. The differences among the variance estimates associated with each volume estimator for a single inventory method (RSRS) are displayed graphically. The bootstrap method of computing the variance estimate is observed to be the only variance computation that consistently mimicked the simulated standard error.

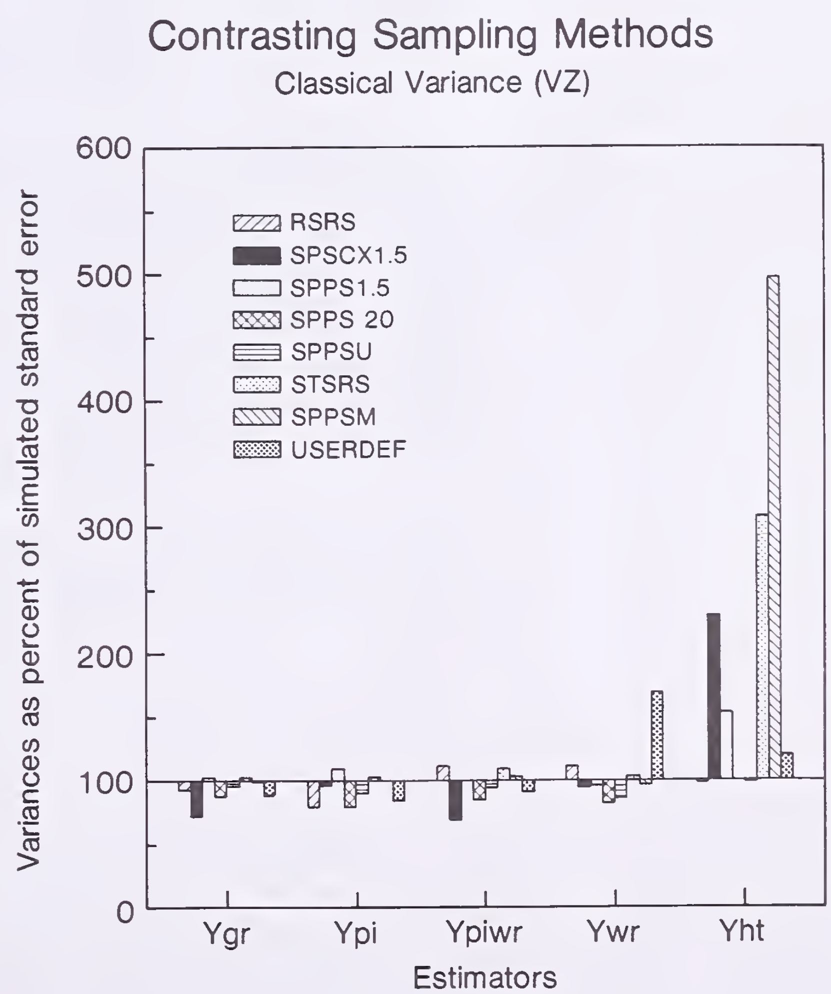


Figure 2. A single variance estimator is graphically displayed for all volume estimators for all inventory methods. The different regression estimators perform better across all inventory methods, but the Horvitz-Thompson estimator fared well for RSRS, STSRS, and SPPS1.5.

Literature Cited

Arvanitis, L.G.; Reich, R.M. 1989. Sampling simulation with a microcomputer. In COENOSES. 4(2):73–80.

Cochran, W.G. 1977. Sampling techniques. 3d ed. New York: Wiley and Sons. 428 p.

Cumberland, W.G.; Royall, R.M. 1981. Prediction models and unequal probability sampling. Journal of Royal Statistical Society B. 43: 353–367.

Efron, B.; Tibshirani, R. 1991. Statistical data analysis in the computer age. Science. 253: 390–395.

Kahaner, D.; Moler, C.; Nash, S. 1988. Numerical methods and software. Englewood Cliffs, NJ: Prentice Hall: 395–397

L'Ecuyer, Pierre. 1988. Efficient and portable combined random number generators. Communications of the ACM. 31(6): 742–749, 774.

L'Ecuyer, Pierre. 1990. Random numbers for simulation. Communications of the ACM. 33(10): 85–97.

Li, H.G.; Brink, G.E.; Schreuder, H.T. In press. An algorithm to set up strata under balanced constraints for optimal sampling strategies. CompStat. In press.

Mackisack, M.S.; Wood, G.B. 1988. FPS-SIM. User guide to forest point sampling simulation package. Draft. Dept. of Forestry, Australian National Univ., Canberra, Australia. 128 p.

Mackisack, M.S.; Wood, G.B. 1990. Simulating the forest and the point-sampling process as an aid in designing forest inventories. Forest Ecology and Management. 36: 79–103.

Ouyang, Z.; Schreuder, H.T.; Li, H. G. In press. Robust regression sampling. Communications in Statistics.

Särndal, C.E. 1980. A two-way classification of regression estimation strategies in probability sampling. Canadian Journal of Statistics 8: 165–177.

Särndal, C.E. 1982. Implications of survey design for generalized regression estimation of linear functions. Journal of Statistical Planning and Information 7: 155–170.

Schreuder, H.T.; Li, H.G.; Wood, G.B. 1990. Model-dependent and design-dependent procedures—a simulation study. Res. Pap. RM-291. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Forest and Range Experiment Station. 19 p.

Schreuder, H.T.; Ouyang, Z. In press. Optimal sampling strategies for weighted linear regression estimation. Canadian Journal of Forest Research.

Snedecor, G.W.; Cochran, W.G. 1967. Statistical methods. 6th ed. Ames, IA: The Iowa State University Press. 593 p.

Zahl, Samuel. 1977. Jackknifing an index of diversity. Ecology. 58: 907–913.

Brink, Glen E.; Schreuder, Hans T.; 1991. ONEPHASE: A Simulation Program to Compare 1-Phase Sampling Strategies. Res. Pap. RM-302. Fort Collins, CO; U.S. Department of Agriculture, Forest Service, Rocky Mountain Forest and Range Experiment Station. 8 p.

ONEPHASE is a computer simulation program primarily intended for use by students in Biometry or Forest Mensuration. Using real or artificial populations, it simulates the results of several inventory sampling techniques. Both volume estimates and variances are generated and the results are displayed for comparison and analysis.

There is no cost for this computer program. However, the requestor must provide a formatted, double-sided, double-density or high-density "floppy" (5 1/4" or 3 1/2") diskette suitable for use in IBM personal computers (PC's) or compatibles, and enclose a self-addressed, postage-paid mailer with suitable protection for the diskette. Execution of ONEPHASE requires an IBM-compatible PC with 300K of available memory. For further information write Multiresource Inventory Techniques Research Work Unit, Rocky Mountain Forest and Range Experiment Station, USDA Forest Service, 240 West Prospect Road, Fort Collins, CO 80526–2098.

Keywords: Computer Simulation, Classroom, Forest Inventory, Biometry, Regression, Horvitz-Thompson, Bootstrap, Jackknife Estimation.

U.S. Department of Agriculture
Forest Service

Rocky Mountain Forest and Range Experiment Station

Rocky Mountains

Southwest

Great Plains

The Rocky Mountain Station is one of eight regional experiment stations, plus the Forest Products Laboratory and the Washington Office Staff, that make up the Forest Service research organization.

RESEARCH FOCUS

Research programs at the Rocky Mountain Station are coordinated with area universities and with other institutions. Many studies are conducted on a cooperative basis to accelerate solutions to problems involving range, water, wildlife and fish habitat, human and community development, timber, recreation, protection, and multiresource evaluation.

RESEARCH LOCATIONS

Research Work Units of the Rocky Mountain Station are operated in cooperation with universities in the following cities:

Albuquerque, New Mexico
Flagstaff, Arizona
Fort Collins, Colorado*
Laramie, Wyoming
Lincoln, Nebraska
Rapid City, South Dakota
Tempe, Arizona

*Station Headquarters: 240 W. Prospect Rd., Fort Collins, CO 80526